MPS350/4111: Assignment 1

You can find formulae for named distributions and conjugate pairs in Appendix A of the lecture notes.

⋆⋆ 1. (a) Let $X \sim \text{Pareto}(2, 1)$. Find the probability density function of $X|_{\{X \geq 5\}}$.

(b) Let (X, Y) be continuous random variables with p.d.f. satisfying

$$f_{X,Y}(x, y) \propto \begin{cases} ye^{-\lambda y}x^{y-1} & \text{if } x \in (0, 1) \text{ and } y \in (0, \infty), \\ 0 & \text{otherwise} \end{cases}$$

i. Find the distribution of Y.

ii. Find the distribution of $X|_{\{Y=y\}}$, for $y \in (0, \infty)$.

⋆ 2. In ancient Rome, the *denarius* was a commonly used coin. In this question we are interested in the weight of a denarius coin, which is variable because they were manufactured by hand. A further complication is that, as Roman craftsmanship gradually improved, smaller and smaller versions of the same coin were made.

An analysis of the production methods has suggested that we should model the weight of a single coin, in grams, as $N(\theta, 0.5^2)$. An archaeological dig finds 3 denarii coins, with weights (in grams)

$$x = (3.68, 3.92, 3.85).$$

The archaeologists believe these coins are from the era 37-14 B.C. shortly after the death of Julius Ceaser (in 44 B.C.). They are certain that Julius Ceaser set the official weight of a denarius at 3.9g, but they don't know how closely this was followed, in practice. After some consultation, we decide on a prior $\Theta \sim N(3.9, 1.2^2)$.

(a) Write down a suitable model family M_θ, which leads to a Bayesian model (X, Θ) applicable to the data x.

(b) Use the Normal-Normal conjugate pair to write down the distribution of the posterior $\Theta|_{\{X=x\}}$.

(c) *This part should be done on a computer and is not for handing in. You may find it helpful to use the code provided in Excercise **2.1** (or **4.1**) of the lecture notes.*

Use R or Python to plot the prior and posterior density functions from (b) on the same axes.

(d) Suppose that, in an alternate reality, we had found the data x but had not been able to discuss the choice of prior with the archaeologists. Instead of the prior above we chose $\Theta \sim N(7.5, 4.5^2)$ – based, say, on the rough weights of various modern coins, which are usually between 3g and 12g. Find the resulting posterior $\Theta|_{\{X=x\}}$ and comment on how this compares to the results in (b).

⋆⋆ 3. In this question we use the parametrization of the Geometric distribution from the reference sheet, that is

$$\mathbb{P}[\text{Geometric}(\theta) = k] = \begin{cases} \theta(1-\theta)^k & \text{for } k \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Let $n \in \mathbb{N}$. Let (X, Θ) be a Bayesian model with model family $M_\theta \sim \text{Geometric}(\theta)^{\otimes n}$, where $\theta \in [0, 1]$, and prior $\Theta \sim \text{Beta}(a, b)$, where $a, b \in (0, \infty)$.

Let $x = (x_1, \ldots, x_n) \in (\mathbb{N}_0)^n$. Show that the posterior distribution is

$$\Theta|_{\{X=x\}} \sim \text{Beta}\left(a + n, b + \sum_{i=1}^{n} x_i\right).$$

*In this question you should justify your calculations **without** using the reference sheet of conjugate pairs.*

(b) The probability of a wild oyster containing a pearl is around one in ten thousand. This probability can be greatly increased in farmed oysters with various techniques, but it is still an uncertain process.

An oyster farmer opens oysters one by one. We use a Geometric(θ) distribution to model the number of times the farmer opens an oyster and does not find a pearl, up to (but not including) the time at which the first pearl is found. We decide to use the prior $\Theta \sim \text{Beta}(1, 10)$. The farmer has repeated this experiment 10 times and obtained the data

$$x = (25,\ 38,\ 11,\ 60,\ 23,\ 29,\ 4,\ 28,\ 61,\ 28). \qquad (\star)$$

For this data $\sum_{i=1}^{10} x_i = 307$.

i. What is the range R and the parameter space Π of the discrete Bayesian model specified above?

ii. How many oysters have been opened, in total, to collect the data $(\star)$?

iii. Use part (a) to find the posterior distribution $\Theta|_{\{X=x\}}$, using the data supplied by the oyster farmer. What is the value of $\mathbb{E}[\Theta|_{\{X=x\}}]$ and how does this compare to our prior?

iv. Suppose that, instead, we choose to model the probability of finding a pearl inside a *single* oyster as Bernoulli(θ).
Let (Y, Θ) be a Bayesian model with model family $\text{Bernoulli}(\theta)^{\otimes 317}$, with the same prior $\Theta \sim \text{Beta}(1, 10)$. Use the reference sheet of conjugate pairs to find the posterior parameters for this model, given the data represented in $(\star)$.

v. You should obtain the same posterior distribution in parts iii and iv. Suggest why has this happened.